1. Introduction

The goal of this project was to create an interactive application that enables users to generate 3D anaglyph images by inserting a segmented person into a stereoscopic image pair. The project integrates deep learning, image processing, and user interface design. Using a pre-trained semantic segmentation model, a person can be extracted from any given image and then inserted into the left and right images of a stereo pair with custom positioning and depth. The left and right views are then merged into a red-cyan anaglyph image suitable for 3D viewing. This project was successfully implemented using PyTorch for model inference and Gradio for the user interface.

2. Methodology

The application was developed using a modular pipeline involving several key components:

1. **Image Segmentation**
   A pre-trained DeepLabV3 model with a ResNet-101 backbone from `torchvision.models.segmentation` was used to extract a mask of the person. The model was chosen for its strong balance between performance and integration simplicity. Its ability to generate dense pixel-level segmentation, along with ready-to-use COCO-trained weights, made it ideal for this application.

2. **Why DeepLabV3?**
   DeepLabV3 was selected because it is accurate, efficient, and tightly integrated with the PyTorch ecosystem. It uses Atrous Spatial Pyramid Pooling (ASPP) for multi-scale context aggregation, which is helpful for segmenting people in complex scenes. Unlike Mask R-CNN or SegFormer, DeepLabV3 requires no external libraries and is fast enough for interactive use. It directly supports isolating the “person” class (ID 15 in COCO), which simplifies the processing pipeline.

3. **Stereo Image Processing**
   A side-by-side stereo image was split into left and right views using PIL. This approach allowed the use of real-world stereo photography while maintaining alignment and resolution consistency.

4. **Depth and Disparity Control**
   Depth perception was controlled by introducing a horizontal offset (disparity) between the person’s position in the left and right images. Greater disparity resulted in the person appearing closer to the viewer, while smaller disparity made them appear further away.

5. **Image Composition**
   The segmented and optionally scaled person image was composited onto both left and right images. The user could interactively adjust the X and Y positions, as well as scale, to place the figure naturally into the scene.

6. **Anaglyph Generation**
   The red channel from the left image was merged with the green and blue channels from the right image to create a red-cyan anaglyph. This format is viewable with 3D glasses and simulates depth perception effectively.

7. **Interactive Interface with Gradio**
   A Gradio app was developed to let users upload images, select the desired depth level, and adjust the position and size of the person. The interface was made user-friendly and dynamically adapted to the uploaded image size.

3. Results

The final application successfully allowed for interactive insertion of a segmented person into a 3D stereo scene and generated high-quality anaglyph images. An example output is shown below:

In this result, the person was placed on top of a building within the cityscape. Using the adjustable controls, their position and scale were fine-tuned for realistic integration. When viewed with red-cyan 3D glasses, the figure appears to stand out clearly from the background due to the applied disparity. The surrounding buildings and vehicles retain their natural stereo depth, contributing to the realism of the scene.

4. Discussion

The project was generally successful, but several insights and challenges emerged during development:

Segmentation worked well in most cases but could produce artifacts near the edges of the person, especially when the background was visually complex. While DeepLabV3 offered reliable masks, some refinement (e.g., feathering or alpha blending) could improve integration quality. Disparity tuning proved to be a sensitive process — even small shifts could significantly change the perceived depth, requiring thoughtful control from the user.

The generated anaglyph image effectively demonstrated stereo depth. The inserted figure displayed consistent left-right alignment, and the red-cyan merge maintained a convincing 3D effect. However, realism could be further improved by adding shadows or better color blending to harmonize the foreground with the background. Also, stereo ghosting was occasionally observed, particularly around high-contrast edges — a known limitation of anaglyph techniques.

Overall, the project validated that an end-to-end 3D insertion and visualization tool can be built using deep learning, stereo imaging, and user-driven interfaces.

5. Conclusion

This lab project successfully combined semantic segmentation, stereo image processing, and UI interaction into a fully functional 3D anaglyph image generator. By leveraging PyTorch and Gradio, the system allowed users to isolate a person, insert them into a stereo scene at adjustable depth, and generate compelling red-cyan 3D outputs.

Future enhancements could include improving segmentation quality, supporting lighting-aware insertion with shadows, handling multiple people or dynamic subjects, and extending the system to support video or animation. The experience provided valuable insights into integrating deep learning models with interactive image editing and gave a deeper understanding of stereo vision principles in practice.